



SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: BILLING-MEDEL, PATRICIA A.
COHEN, MAURICE
COLPITTS, TRACEY L.
FRIEDMAN, PAULA N.
HAYDEN, MARK
KLASS, MICHAEL R.
ROBERTS-RAPP, LISA
RUSSELL, JOHN C.
STROUPE, STEPHEN D.

(ii) TITLE OF THE INVENTION: REAGENTS AND METHODS FOR THE
USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL
TRACT

(iii) NUMBER OF SEQUENCES: 51

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Abbott Laboratories
- (B) STREET: 100 Abbott Park Road
- (C) CITY: Abbott Park
- (D) STATE: IL
- (E) COUNTRY: USA
- (F) ZIP: 60064-3500

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: 08/828,856
- (B) FILING DATE: 31-MAR-1997

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Becker, Cheryl L.
- (B) REGISTRATION NUMBER: 35,441.
- (C) REFERENCE/DOCKET NUMBER: 6068.US.P1

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 847/935-1729
- (B) TELEFAX: 847/938-2623
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCCAGGAATA ACTAGAGAGG AACAAATGGGG TTATTCAAGAG GTTTTGTGTT CCTCTTAGTT	60
CTGTGCCTGC TGACCAAGTC AAATACTTC TTCATTAAGC TGAATAATAA TGGCTTGAA	120
GATATTGTCA TTGTTATAGA TCCTAGTGTG CCAGAAGATG AAAAATAAT TGAACAAATA	180
GAGGATATGG TGACTACAGC TTCTACGTAC CTGTTGAAG CCACAGAAAA AAGATTTTT	240
T	241

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTAGAGAGGA ACAATGGGGT TATTCAAGAGG TTTTGTGTTT CTCTTAGTTC TGTGCCTGCT	60
GCACCAAGTCA AATACTTCCT TCATTAAGCT GAATAATAAT GGCTTGAAAG ATATTGTCAT	120
TGTTATAGAT CCTAGTGTGC CAGAAGATGA AAAAATAATT GAACAAATAG AGGATATGGT	180
GACTACAGCT TCTACGTACC TGTTGAAGC CACAGAAAA	219

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 3
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTNTGTAACG AAAAACCCA TAATCAAGAA GCTCCAAGCC TACAAAACAT AAAGTGCAT	60
TTTAGAAGTA CATGGGAGGT GATTAGCAAT TCTGAGGATT TTAAAAACAC CATAACCATG	120
GTGACACCCAC CTCCTCCACC TGTCTCTCA TTGCTGAAGA TCAGTCAAAG AATTGTGTGC	180
TTAGTTCTTG ATAAGTCTGG AAGCATGGGG GGTAAAGGACC GCCTAAATCG A	231

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 237 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGGGGGGTAA GGACCGCCTA AATCGAATGA ATCAAGCAGC AAAACATTTC CTGCTGCAGA	60
CTGTTGAAAA TGGATCCTGG GTGGGGATGG TTCACTTGA TAGTACTGCC ACTATTGTAA	120
ATAAGCTAAT CAAATAAAA AGCAGTGATG AAAGAAACAC ACTCATGGCA GGATTACCTA	180
CATATCCTCT GGGAGGAAC TCCATCTGCT CTGGAATTAA ATATGCATTT CAGGTGA	237

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 216 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTTCCATCTG CTCTGGATT AAATATGCAT TTCAGGTGAT TGGAGAGCTA CATTCCAAC	60
TCGATGGATC CGAAGTACTG CTGCTGACTG ATGGGGAGGA TAACACTGCA AGTTCTTGTAA	120
TTGATGAAGT GAAACAAAGT GGGCCATTG TTCATTTTAT TGCTTTGGGA AGAGCTGCTG	180
ATGAAGCAGT ATAGAGAGATG AGCAAGATAA CAGGAG	216

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 24
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATTGATAGT ACAGTGGGAA AGGNACGTT CTTTCTCATC ACATGGAACA GTCTGCCTCC	60
CAGTATTCT CTCTGGGATC CCAGTGGAAC ATAATGGAA AATTTCACAG TGGATGCAAC	120
TTCCAAAATG GCCTATCTCA GTATTCAGG AACTGCAAAG GTGGGCACCTT GGGCATACAA	180
TCTTCAAGCC AAAGCGAACCC	201

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 241 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCAAATTCTT CTGTGCCTCC AATCACAGTG AATGCTAAAA TGAATAAGGA CGTAAACAGT	60
TTCCCCAGCC CAATGATTGT TTACGCAGAA ATTCTACAAG GATATGTACC TGTTCTTGGAA	120
GCCAATGTGA CTGCTTCAT TGAATCACAG AATGGACATA CAGAAGTTTT GGAACCTTTTG	180
GATAATGGTG CAGGCGCTGA TTCTTCAAG AATGATGGAG TCTACTCCAG GTATTTTACA	240
G	241

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTGCAGGCCG	TGATTCTTC	AAGAATGATG	GAGTCACTC	CAGGTATTT	ACAGCATATA	60
CAGAAAATGG	CAGATATAGC	TTAAAAGTTC	GGGCTCATGG	AGGAGCAAAC	ACTGCCAGGC	120
TAAAATTACG	GCCTCCACTG	AATAGAGCCG	CGTACATACC	AGGCTGGTA	GTGAACGGGG	180
AAATTGAAGC	AAACCCGCCA	AGACCTGAAA	TTGATGAGGA	TACTCAGACC	ACCTTGGAGG	240
AT						242

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 233 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCCGCCAAGA	CCTGAAATTG	ATGAGGATAC	TCAGACCACC	TTGGAGGATT	TCAGCCGAAC	60
AGCATCCGGA	GGTGCAATTG	TGGTATCACA	AGTCCAAGC	CTTCCCTTGC	CTGACCAATA	120
CCCACCAAGT	CAAATCACAG	ACCTTGATGC	CACAGTTCAT	GAGGATAAGA	TTATTCTTAC	180
ATGGACAGCA	CCAGGAGATA	ATTTTGATGT	TGGAAAAGTT	CAACGTTATA	TCA	233

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 313 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ix) FEATURE:
 (A) NAME/KEY: base_polymorphism
 (B) LOCATION: 22
 (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:
 (A) NAME/KEY: base_polymorphism
 (B) LOCATION: 44
 (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCAATACCCA	CCAAGTCAAA	TNACAGACCT	TGATGCCACA	GTTNATGAGG	ATAAGATTAT	60
TCTTACATGG	ACAGCACCAG	GAGATAATT	TGATGTTGGA	AAAGTTCAAC	GTTATATCAT	120
AAGAATAAGT	GCAAGTATTC	TTGATCTAAG	AGACAGTTT	GATGATGCTC	TTCAAGTAAA	180
TACTACTGAT	CTGTCACCAA	AGGAGGCCA	CTCCAAGGAA	AGCTTGCAT	TTAAACCAGA	240
AAATATCTCA	GAAGAAAATG	CAACCCACAT	ATTATTGCC	ATTAAGGAGTA	TAGATAAAC	300
ATTTGGCATC	AAA					313

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 242 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAGTATTCTT GATCTAAGAG ACAGTTTGA TGATGCTCTT CAAGTAAATA CTACTGATCT	60
GTCACCAAAG GAGGCCAACT CCAAGGAAAG CTTTGCATT AAACCAGAAA ATATCTCAGA	120
AGAAAATGCA ACCCACATAT TTATTGCCAT TAAAAGTATA GATAAAAGCA ATTTGACATC	180
AAAAGTATCC AACATTGCAC AAGTAACCTT GTTTATCCCT CAAGCAAATC CTGATGACAT	240
TG	242

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 208 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 2
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 4
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ANANAATGCA ACCCACATAT TTATTGCCAT TAAAAGTATA GATAAAAGCA ATTTGACATC	60
AAAAGTATCC AACATTGCAC AAGTAACCTT GTTTATCCCT CAAGCAAATC CTGATGACAT	120
TGATCCTACT CCTACTCCTA CTCCTACTCC TGATAAAAGT CATAATTCTG GAGTTAACAT	180
TTCTACGCTG GTATTGTCTG TGATTGGG	208

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTCCTACTCC TACTCCTGAT AAAAGTCATA ATTCTGGAGT TAATATTCT ACGCTGGTAT	60
TGTCTGTGAT TGGGTCTGTT GTAATTGTTA ACTTTATTTT AAGTACCAACC ATTTGAACCT	120
TAACGAAGAA AAAAATCTTC AAGTAGACCT AGAAGAGAGT TTTAAAAAAC AAAACAATGT	180
AAGTAAAGGA TATTCTGAA T	201

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 301 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: base_polymorphism
- (B) LOCATION: 111
- (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:
 (A) NAME/KEY: base_polymorphism
 (B) LOCATION: 244
 (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(ix) FEATURE:
 (A) NAME/KEY: base_polymorphism
 (B) LOCATION: 284
 (D) OTHER INFORMATION: /note= "'N' represents an A or G or T or C polymorphism at this position"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCTGTTGTA	TTGTTAAC	TTTAAAGT	ACCACCA	TTT	GAACCTTA	AC	GAAGAAAAAA		60
ATCTTCAGT	AGACCTAGA	GAGAGTTT	AAA	ACAAAAA	CAATGTAA	GT	NAAGGATATT		120
TCTGAATCTT	AAAATT	CATC	CCATGTGT	GA	TCATAAA	ACTC	ATAAAAA	TTTAAGATG	180
TCGGAAAAGG	ATAC	TTTGAT	TAA	AAAAAA	CACTCATG	GA	TATGAAAAA	CTGTCAAGAT	240
TAANATTAA	TAGTT	CATT	TATT	GTTAT	TTTATTG	TG	AGANATAGTG	ATGAACAAAG	300
A									301

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 229 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GA	AAAAAAATC	TTCAAGTAGA	CCTAGAAGAG	AGTTT	AAAAA	ACACAA	TGTAAGTAAA		60
GG	ATTTCT	GAATCTTAA	ATTCA	CCC	TG	TGATCA	TAAACTCATA	AAA	120
TA	AGATGT	CG	AAAAGGATA	CTTGATTAA	ATAAA	ACAC	TCATGGATAT	GTAAA	180
TC	AAAGATTAA	AATT	AA	TTAATAG	TT	TGATT	TTT	TTGTAAG	229

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3043 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTAGAGAGGA	ACAATGGG	TATT	CAGAGG	TTT	GTTT	TC	CTTAGTTC	TGTGCCTG	60
GCACCA	GTCA	AAT	ACTT	C	TCA	TAAAG	CT	GG	120
TGTT	TAT	AGAT	C	TG	GA	AAATA	AA	TTG	180
GACTAC	AGCT	TCTAC	GT	TTG	GA	AAAATA	AA	ATG	240
ATCT	TAT	ATTA	TT	GA	AG	AAACAA	AC	GT	300
AAAC	CATA	AA	CTG	GA	AG	AACTCT	TC	AA	360
ATAC	CCA	AT	GG	AA	GG	GGCG	CC	AC	420
CCTT	CTACT	AA	GTG	AA	AGG	GA	CC	CT	480
GTGG	GCT	AA	GGT	AA	GG	GGCA	CC	GT	540
TGCT	TA	AA	GGT	AA	GG	AACT	CT	TC	600
AGTT	TATA	AA	GGG	AA	GG	TGAG	TT	AC	660
AAA	ACTGT	AA	GGG	AA	GG	TACAA	TT	AC	720
CATA	ATG	AA	GGT	AA	GG	AAAC	TT	AC	780
TCAAGA	GAGCT	AA	GGT	AA	GG	ACAG	TT	AC	840
TAGCAATT	TCT	AA	GGG	AA	GG	AAAC	AC	CT	900
		AA	GGG	AA	GG	CCATGG	CC	AC	
		AA	GGG	AA	GG	GGT	CC	AC	

CTTCTCATTG	CTGAAGATCA	GTC	AAAGAAT	TGTGTGCTTA	GTTCTTGATA	AGTCTGGAA	960
CATGGGGGGT	AAGGACCGCC	TAATCGA	AT	GAATCAAGCA	GCAAACATT	TCCTGTCGA	1020
GA	CTGTTGAA	AATGGATCCT	GGGTGGGGAT	GGTCACTT	GATAGTACTG	CCACTATTGT	1080
AAATAAGCTA	ATCCAATAA	AAAGCAGTGA	TGAAAGAAC	ACACTCATGG	CAGGATTACC	1140	
TACATATCCT	CTGGGAGGAA	CTTCCATCTG	CTCTGGAATT	AAATATGCAT	TTCAGGTGAT	1200	
TGGAGAGCTA	CATTCCCAC	TCGATGGATC	CGAAGTACTG	CTGCTGACTG	ATGGGGAGGA	1260	
TAACACTGCA	AGTTCTTGT	TTGATGAAGT	GAAACAAAGT	GGGCCATTG	TTCATTTTAT	1320	
TGCTTGGGA	AGAGCTGCTG	ATGAAGCAGT	ATAGAGATG	AGCAAGATAA	CAGGAGGAAG	1380	
TCATTTTAT	GTTTCAGATG	AAGCTCAGAA	CAATGGCCTC	ATTGATGCTT	TTGGGGCTCT	1440	
TACATCAGGA	AATACTGATC	TCTCCAGAA	GTCCCTTCAG	CTCGAAAGTA	AGGGATTAAC	1500	
ACTGAATAGT	AATGCCTGGA	TGAACGACAC	TGTCTAATT	GATAGTACAG	TGGGAAAGGA	1560	
CACGTTCTT	CTCATCACAT	GGAACAGTCT	CCCTCCAGT	ATTTCTCT	GGGATCCCAG	1620	
TGGAACAATA	ATGGAAAATT	TCACAGTGG	TGCAACTTC	AAAATGGCCT	ATCTCAGTAT	1680	
TCCAGGAAC	GCAAAGGTGG	GCAC	TTGGGC	ATACAATCTT	CAAGCCAAAG	1740	
AACATTAAC	ATTACAGTAA	CTTCTCGAGC	AGCAAATTCT	TCTGTGCC	CAATCACAGT	1800	
GAATGCTAA	ATGAATAAGG	ACGTAAACAG	TTTCCCAGC	CCAATGATTG	TTTACGCAGA	1860	
AATTCTACAA	GGATATGTAC	CTGTTCTT	AGCCAATGTG	ACTGCTTCA	TTGAATCACA	1920	
GAATGGACAT	ACAGAAGTTT	TGGAAC	TTTT	GGATAATGGT	GCAGGGCCTG	1980	
GAATGATGGA	GTC	TACTCCA	GGTATTTAC	AGCATATACA	GAAATGGCA	2040	
AAAAGTTCGG	GTCATGGAG	GAGCAACAC	TGCCAGGCTA	AAATTACGGC	CTCCACTGAA	2100	
TAGAGCCCG	TACATACCG	GCTGGTAGT	GAACGGGGAA	ATTGAAGCAA	ACCCGCCAAG	2160	
ACCTGAAATT	GATGAGGATA	CTCAGACCAC	CTTGGAGGAT	TTCAGCCGAA	CAGCATCCGG	2220	
AGGTGCA	TTGGTATCAC	AAGTCCAA	GGCTCCCTTG	CCTGACCAAT	ACCCACCAAG	2280	
TCAAATCACA	GACCTTGATG	CCACAGTTCA	TGAGGATAAG	ATTATTCTTA	CATGGACAGC	2340	
ACCAGGAGAT	AATTTTGATG	TTGGAAAAGT	TCAACGTTAT	ATCATAAGAA	TAAGTGCAG	2400	
TATTCTGAT	CTAAGAGACA	GTTTGATGA	TGCTCTTCA	GTAAATACTA	CTGATCTGTC	2460	
ACCAAAGGAG	GCCAAC	TCCA	AGGAAAGCTT	TGCATTTAA	CCAGAAAATA	2520	
AAATGCAACC	CACATATTTA	TTGCCATTAA	AAGTATAGAT	AAAAGCAATT	TGACATCAA	2580	
AGTATCCAAC	ATTGCACAAG	TAAC	TTGTT	TATCCCTCA	GCAAATCCTG	2640	
TCCTACTCCT	ACTCCTACTC	CTACTCCTGA	AAAAGTCAT	AATTCTGGAG	TTAATATTC	2700	
TACGCTGGT	TTGCTGTG	TTGGGCTCTG	TGTAATTGTT	AACTTTATT	TAAGTAC	2760	
CATTTGAACC	TTAACGAGA	AAAAAACTCTT	CAAGTAGAC	TAGAAGAGAG	TTTTAAAAAA	2820	
CAAAACAATG	TAAGTAAAGG	ATATTC	TGTA	ATCTAAAAT	TCATCCCAGT	2880	
AACTCATAA	AATAATT	TTAGT	CGGA	AAAGGATACT	TTGATTAAT	AAAACACTC	2940
ATGGATATGT	AAAAACTGTC	AAGAT	TTAATAGTT	TCATTTATT	GTTATTTAT	3000	
TTGTAAGAAA	TAGT	GATGAA	CAAAGATCCT	TTTC	CATACT	3043	

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1399 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCAAATTCTT	CTGTGCC	TCC	AATCACAGT	AATGCTAAA	TGAATAAGGA	CGTAAACAGT	60
TTCCCCAGCC	CAATGATTG	T	TACG	CAGAA	ATTCTACAA	GATATG	120
GCCAATG	CTG	TTT	CAT	TGAATC	ACAG	ATGGAC	180
GATAATGGT	CAG	CGC	GCTG	TTCTT	CAAG	ATGGAG	240
GCATATACAG	AAA	ATGGC	AG	ATAGCTT	AAAGT	TCGGGG	300
GCCAGGCTAA	AAT	ACGGCC	TCC	ACTGA	AGAGCC	CGT	360
AACGGGGAAA	TTG	AGC	AAA	CCG	CCAGA	CCTGAA	420
TTGGAGGATT	TCAG	CCG	AA	GCAT	CCG	GGTCA	480
CTTCCCTG	CTG	ACCA	AA	CC	ACCA	AGTCA	540
GAGGATAAGA	TTAT	TCT	TA	GGAC	AGA	GGAGA	600
CAACGTTATA	TCAT	AAGA	A	AAGT	CAAGT	TTCTGATG	660
GCTCTTCAAG	TAAT	ACTAC	T	GATCTG	CCAA	GGAGGAGG	720
GCATTTAAC	CAG	AAA	AT	TCAGA	AA	ATGCAAC	780
AGTATAGATA	AA	AGCA	TTT	AAC	AT	TGACACA	840
ATCCCTCAAG	CAA	ATC	CTG	TGAC	ATTG	TCTACTC	900
AAAAGTCATA	ATT	TGG	AGT	ACG	TGGT	GTAT	960
GTAATTGTTA	ACT	TTA	TTT	AAGT	ACCA	CT	1020

AAGTAGACCT	AGAAGAGAGT	TTTAAAAAAC	AAAACAATGT	AAGTAAAGGA	TATTCTGAA	1080
TCTTAAAATT	CATCCCATGT	GTGATCATAA	ACTCATAAAA	ATAATTAA	GATGTCGGAA	1140
AAGGATACTT	TGATTAATAA	AAAACACTCA	TGGATATGTA	AAAACTGTCA	AGATTAAAAT	1200
TTAATAGTTT	CATTATTG	TTATTTATT	TGTAAGAAAT	AGTGATGAAAC	AAAGATCCTT	1260
TTTCATACTG	ATACCTGGTT	GTATATTATT	TGATGCAACA	GTTTCTGAA	ATGATATTTC	1320
AAATTGCATC	AAGAAATTAA	AATCATCTAT	CTGAGTAGTC	AAAATACAAG	TAAGGAGAG	1380
CAAATAAACAA	ACATTGGAA					1399

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCCAGGAATA	ACTAGAGAGG	AACAATGGGG	TTATTCAAGAG	GTTTGT	TTT CCTCTTAGTT	60
CTGTGCCTGC	TGCACCAAGTC	AAATACTTCC	TTCATTAAGC	TGAATAATAA	TGGCTTTGAA	120
GATATTGTCA	TTGTTATAGA	TCCTAGTGTG	CCAGAAGATG	AAAAAATAAT	TGAACAAATA	180
GAGGATATGG	TGACTACAGC	TTCTACGTAC	CTGTTGAG	CCACAGAAAA	AAGATTTTTT	240
TTCAAAATG	TATCTATATT	AATTCTGAG	AATTGGAAGG	AAAATCCTCA	GTACAAAAGG	300
CCAAAACATG	AAAACCATAA	ACATGCTGAT	GTTATAGTTG	CACCACCTAC	ACTCCCAGGT	360
AGAGATGAAC	CATACACCAA	GCAGTTCA	GAATGTGGAG	AGAAAGGCAGA	ATACATTCAC	420
TTCACCCCTG	ACCTTCTACT	TGAAAAAAA	CAAATGAAT	ATGGACCACC	AGGCAAACGT	480
TTTGTCCATG	AGTGGGCTCA	CCTCCGGTGG	GGAGTGT	ATGAGTACAA	TGAAGATCAG	540
CCTTTCTACC	GTGCTAAGTC	AAAAAAATC	GAAGCAACAA	GGTGTCCGC	AGGTATCTCT	600
GTAGAAATA	GAGTTTATAA	GTGTCAAGGA	GGCAGCTGTC	TTAGTAGAGC	ATGCAGAATT	660
GATTCTACAA	CAAACACTGTA	TGGAAAAGAT	TGTCATTCT	TTCCCTGATAA	AGTACAAACA	720
AAAAAAAGCAT	CCATAATGTT	TATGCAAAGT	ATTGATTCTG	TTGTTGAATT	TTGTAACGAA	780
AAAACCCATA	ATCAAGAAGC	TCCAAGCTA	CAAACATAA	AGTCAATT	TAGAAGTACA	840
TGGGAGGTGA	TTAGCAATT	TGAGGATTT	AAAAACACCA	TACCCATGGT	GACACCACCT	900
CCTCCACCTG	TCTTCTCATT	GCTGAAGATC	AGTCAAAGAA	TTGTTGCTT	AGTCTTGAT	960
AAGTCTGAA	GCATGGGGGG	TAAGGACCGC	CTAAATCGAA	TGAATCAAGC	AGCAAAACAT	1020
TTCCTGCTGC	AGACTGTTGA	AAATGGATCC	TGGGTGGGG	TGGTCACTT	TGATAGTACT	1080
GCCACTATTG	TAAATAAGCT	AATCCAATAA	AAAAGCAGTG	ATGAAAGAAA	CACACTCATG	1140
GCAGGATTAC	CTACATATCC	TCTGGGAGGA	ACTTCCATCT	GCTCTGGAA	TAAATATGCA	1200
TTTCAGGTGA	TTGGAGAGCT	ACATTCCAA	CTCGATGGAT	CCGAAGTACT	GCTGCTGACT	1260
GATGGGGAGG	ATAACACTGC	AAAGTTCTGT	ATTGATGAG	TGAAACAAAG	TGGGCCATT	1320
GTTCATTTA	TTGCTTGGG	AAGAGCTGCT	GATGAAGCAG	TAATAGAGAT	GAGCAAGATA	1380
ACAGGGAGGA	GTCATTTTA	TGTTTCAGAT	GAAGCTCAGA	ACAATGGCT	CATTGATGCT	1440
TTTGGGCTC	TTACATCAGG	AAATACTGAT	CTCTCCAGA	AGTCCCTCA	GCTCGAAAGT	1500
AAGGGATTAA	CACTGAATAG	TAATGCCCTGG	ATGAACGACA	CTGTCAATA	TGATAGTACA	1560
GTGGGAAAGG	ACACGTTCTT	TCTCATCACA	TGGAACAGTC	TGCCTCCAG	TATTTCTCTC	1620
TGGGATCCA	GTGGAACAAT	AATGGAAAAT	TTCACAGTGG	ATGCAACTTC	CAAATGGCC	1680
TATCTCAGTA	TTCCAGGAAC	TGCAAAGGTG	GGCACCTTGGG	CATACAATCT	TCAAGCCAAA	1740
GGGAACCCAG	AAACATTAAC	TATTACAGTA	ACTTCTCGAG	CAGCAAATTC	TTCTGTGCCT	1800
CCAATCACAG	TGAATGCTAA	AATGAATAAG	GACGTAAACA	GTTTCCCCAG	CCCAATGATT	1860
GTTCACGCAG	AAATTCTACA	AGGATATGTA	CCTGTTCTTG	GAGCCAATGT	GACTGTTTC	1920
ATTGAATCAC	AGAATGGACA	TACAGAAGTT	TTGGAAC	TGGATAATGG	TGCAGGCCT	1980
GATTCTTCA	AGAATGATGG	AGTCTACTCC	AGGTATTTA	CAGCATATAC	AGAAAATGGC	2040
AGATATAGCT	AAAAGTTCG	GGCTCATGGA	GGAGCAAACA	CTGCCAGGCT	AAAATTACGG	2100
CCTCCACTGA	ATACAGCCGC	GTACATACCA	GGCTGGGTAG	TGAACGGGA	AATTGAAGCA	2160
AACCCGCCAA	GACCTGAAAT	TGATGAGGAT	ACTCAGACCA	CCTTGGAGGA	TTTCAGCCGA	2220
ACAGCATCCG	GAGGTGCATT	TGTGGTATCA	CAAGTCCCAA	GCCTCCCTT	GCCTGACCAA	2280
TACCCACCAA	GTCAAATCAC	AGACCTTGT	GCCACAGTTC	ATGAGGATAA	GATTATTCTT	2340
ACATGGACAG	CACCAAGGAGA	TAATTTGAT	GTTGGAAAAG	TTCAACGTTA	TATCATAAGA	2400
ATAAGTGCAA	GTATTCTGA	TCTAAGAGAC	AGTTTGATG	ATGCTCTCA	AGTAAATACT	2460
ACTGATCTGT	CACCAAAGGA	GGCCAAC	AAGGAAAGCT	TTGCATTAA	ACCAAGAAAAT	2520
ATCTCAGAAC	AAAATGCAAC	CCACATATT	ATTGCCATT	AAAGTATAGA	AAAAAGCAAT	2580
TTGACATCAA	AAGTATCCAA	CATTGCAACAA	GTAAC	TTATCCCTCA	AGCAAATCCT	2640
GATGACATTG	ATCCTACTCC	TACTCCTACT	CCTACTCCTG	ATAAAAGTC	TAATTCTGGA	2700
GTAAATATT	CTACGCTGGT	ATTGTCTGTG	ATTGGGTCTG	TTGTAATTGT	TAACCTTATT	2760
TTAAGTACCA	CCATTGAAC	CTTAACGAAG	AAAAAAATCT	TCAAGTAGAC	CTAGAAGAGA	2820

GTTTAAAAAA ACAAAACAAT GTAAGTAAAG GATATTCTG AATCTTAAAA	TTCATCCCAT	2880
GTGTGATCAT AAACTCATAA AAATAATTTT AAGATGTCGG AAAAGGATAC	TTTGATTAAGA	2940
TAAAAACACT CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT	TTCATTTATT	3000
TGTTATTTA TTTGTAAGAA ATAGTGATGA ACAAAAGATCC	TTTTTCATAC	3060
TTGTATATTA TTTGATGCAA CAGTTTCTG AAATGATATT TCAAATTGCA	TGATACCTGG	3120
AAAATCATCT ATCTGAGTAG TCAAAATACA AGTAAAGGAG AGCAAATAAA	TCAAGAAATT	3180
A	CAACATTGG	3181

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AGCTCGGAAT TCCGAGCTTG GATCCTCTAG AGCGGCCGCC	GACTAGTGAG	CTCGTCGACC	60
CGGGAATT			68

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AATTAATTCC CGGGTCGACG AGCTCACTAG TCGGCGGCCG	CTCTAGAGGA	TCCAAGCTCG	60
GAATTCCG			68

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGCGGATAAC AATTCACAC AGGA		24
---------------------------	--	----

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGTAAAACGA CGGCCAGT		18
---------------------	--	----

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTGCCAGGCT AAAATTACGG

20

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATCACAGACC TTGATGCCAC

20

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCTGGTATTG TCTGTGATTG GGTC

24

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CATCAGGATT TGCTTGAGGG

20

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TATTGGTCAG GCAAGGGAAG

20

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

G TGTTTGCTC CTCCATGAGC

20

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

C AAGTAGAAG GTCAGGGGTG

20

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

A TAA GTGTCA AGGAGGCAGC

20

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

G CAGACTGTT CCATGTGATG

20

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

A TGTACCTGT TCTTGGAGCC

20

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACGTACCTGT TTGAAGCCAC

20

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGTAAGGACC GCCTAAATCG

20

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAAGTGAAAC AAAGTGGGGC

20

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTATCCTCCC CATCAGTCAG

20

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TCGATTTAGG CGGTCTTAC

20

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TGTGGCTTCA AACAGGTACG

20

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGGTAAGGAC CGCCTAAATC GAATG

25

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GAGCCCCAAA AGCATCAATG AGG

23

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 917 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met	Gly	Leu	Phe	Arg	Gly	Phe	Val	Phe	Leu	Leu	Val	Leu	Cys	Leu	Leu
1					5				10				15		
His	Gln	Ser	Asn	Thr	Ser	Phe	Ile	Lys	Leu	Asn	Asn	Asn	Gly	Phe	Glu
					20			25					30		
Asp	Ile	Val	Ile	Val	Ile	Asp	Pro	Ser	Val	Pro	Glu	Asp	Glu	Lys	Ile
					35			40			45				
Ile	Glu	Gln	Ile	Glu	Asp	Met	Val	Thr	Thr	Ala	Ser	Thr	Tyr	Leu	Phe
					50			55			60				
Glu	Ala	Thr	Glu	Lys	Arg	Phe	Phe	Lys	Asn	Val	Ser	Ile	Leu	Ile	
					65			70			75		80		
Pro	Glu	Asn	Trp	Lys	Glu	Asn	Pro	Gln	Tyr	Lys	Arg	Pro	Lys	His	Glu
					85			90			95				
Asn	His	His	Ala	Asp	Val	Ile	Val	Ala	Pro	Pro	Thr	Leu	Pro	Gly	
					100			105			110				
Arg	Asp	Glu	Pro	Tyr	Thr	Lys	Gln	Phe	Thr	Glu	Cys	Gly	Glu	Lys	Gly
					115			120			125				
Glu	Tyr	Ile	His	Phe	Thr	Pro	Asp	Leu	Leu	Leu	Glu	Lys	Lys	Gln	Asn
					130			135			140				
Glu	Tyr	Gly	Pro	Pro	Gly	Lys	Leu	Phe	Val	His	Glu	Trp	Ala	His	Leu
					145			150			155		160		

Arg Trp Gly Val Phe Asp Glu Tyr Asn Glu Asp Gln Pro Phe Tyr Arg
 165 170 175
 Ala Lys Ser Lys Lys Ile Glu Ala Thr Arg Cys Ser Ala Gly Ile Ser
 180 185 190
 Gly Arg Asn Arg Val Tyr Lys Cys Gln Gly Ser Cys Leu Ser Arg
 195 200 205
 Ala Cys Arg Ile Asp Ser Thr Thr Lys Leu Tyr Gly Lys Asp Cys Gln
 210 215 220
 Phe Phe Pro Asp Lys Val Gln Thr Glu Lys Ala Ser Ile Met Phe Met
 225 230 235 240
 Gln Ser Ile Asp Ser Val Val Glu Phe Cys Asn Glu Lys Thr His Asn
 245 250 255
 Gln Glu Ala Pro Ser Leu Gln Asn Ile Lys Cys Asn Phe Arg Ser Thr
 260 265 270
 Trp Glu Val Ile Ser Asn Ser Glu Asp Phe Lys Asn Thr Ile Pro Met
 275 280 285
 Val Thr Pro Pro Pro Pro Val Phe Ser Leu Leu Lys Ile Ser Gln
 290 295 300
 Arg Ile Val Cys Leu Val Leu Asp Lys Ser Gly Ser Met Gly Gly Lys
 305 310 315 320
 Asp Arg Leu Asn Arg Met Asn Gln Ala Ala Lys His Phe Leu Leu Gln
 325 330 335
 Thr Val Glu Asn Gly Ser Trp Val Gly Met Val His Phe Asp Ser Thr
 340 345 350
 Ala Thr Ile Val Asn Lys Leu Ile Gln Ile Lys Ser Ser Asp Glu Arg
 355 360 365
 Asn Thr Leu Met Ala Gly Leu Pro Thr Tyr Pro Leu Gly Gly Thr Ser
 370 375 380
 Ile Cys Ser Gly Ile Lys Tyr Ala Phe Gln Val Ile Gly Glu Leu His
 385 390 395 400
 Ser Gln Leu Asp Gly Ser Glu Val Leu Leu Leu Thr Asp Gly Glu Asp
 405 410 415
 Asn Thr Ala Ser Ser Cys Ile Asp Glu Val Lys Gln Ser Gly Ala Ile
 420 425 430
 Val His Phe Ile Ala Leu Gly Arg Ala Ala Asp Glu Ala Val Ile Glu
 435 440 445
 Met Ser Lys Ile Thr Gly Gly Ser His Phe Tyr Val Ser Asp Glu Ala
 450 455 460
 Gln Asn Asn Gly Leu Ile Asp Ala Phe Gly Ala Leu Thr Ser Gly Asn
 465 470 475 480
 Thr Asp Leu Ser Gln Lys Ser Leu Gln Leu Glu Ser Lys Gly Leu Thr
 485 490 495
 Leu Asn Ser Asn Ala Trp Met Asn Asp Thr Val Ile Ile Asp Ser Thr
 500 505 510
 Val Gly Lys Asp Thr Phe Phe Leu Ile Thr Trp Asn Ser Leu Pro Pro
 515 520 525
 Ser Ile Ser Leu Trp Asp Pro Ser Gly Thr Ile Met Glu Asn Phe Thr
 530 535 540
 Val Asp Ala Thr Ser Lys Met Ala Tyr Leu Ser Ile Pro Gly Thr Ala
 545 550 555 560
 Lys Val Gly Thr Trp Ala Tyr Asn Leu Gln Ala Lys Ala Asn Pro Glu
 565 570 575
 Thr Leu Thr Ile Thr Val Thr Ser Arg Ala Ala Asn Ser Ser Val Pro
 580 585 590
 Pro Ile Thr Val Asn Ala Lys Met Asn Lys Asp Val Asn Ser Phe Pro
 595 600 605
 Ser Pro Met Ile Val Tyr Ala Glu Ile Leu Gln Gly Tyr Val Pro Val
 610 615 620
 Leu Gly Ala Asn Val Thr Ala Phe Ile Glu Ser Gln Asn Gly His Thr
 625 630 635 640
 Glu Val Leu Glu Leu Leu Asp Asn Gly Ala Gly Ala Asp Ser Phe Lys
 645 650 655
 Asn Asp Gly Val Tyr Ser Arg Tyr Phe Thr Ala Tyr Thr Glu Asn Gly
 660 665 670
 Arg Tyr Ser Leu Lys Val Arg Ala His Gly Gly Ala Asn Thr Ala Arg
 675 680 685

Leu Lys Leu Arg Pro Pro Leu Asn Arg Ala Ala Tyr Ile Pro Gly Trp
 690 695 700
 Val Val Asn Gly Glu Ile Glu Ala Asn Pro Pro Arg Pro Glu Ile Asp
 705 710 715 720
 Glu Asp Thr Gln Thr Thr Leu Glu Asp Phe Ser Arg Thr Ala Ser Gly
 725 730 735
 Gly Ala Phe Val Val Ser Gln Val Pro Ser Leu Pro Leu Pro Asp Gln
 740 745 750
 Tyr Pro Pro Ser Gln Ile Thr Asp Leu Asp Ala Thr Val His Glu Asp
 755 760 765
 Lys Ile Ile Leu Thr Trp Thr Ala Pro Gly Asp Asn Phe Asp Val Gly
 770 775 780
 Lys Val Gln Arg Tyr Ile Ile Arg Ile Ser Ala Ser Ile Leu Asp Leu
 785 790 795 800
 Arg Asp Ser Phe Asp Asp Ala Leu Gln Val Asn Thr Thr Asp Leu Ser
 805 810 815
 Pro Lys Glu Ala Asn Ser Lys Glu Ser Phe Ala Phe Lys Pro Glu Asn
 820 825 830
 Ile Ser Glu Glu Asn Ala Thr His Ile Phe Ile Ala Ile Lys Ser Ile
 835 840 845
 Asp Lys Ser Asn Leu Thr Ser Lys Val Ser Asn Ile Ala Gln Val Thr
 850 855 860
 Leu Phe Ile Pro Gln Ala Asn Pro Asp Asp Ile Asp Pro Thr Pro Thr
 865 870 875 880
 Pro Thr Pro Thr Pro Asp Lys Ser His Asn Ser Gly Val Asn Ile Ser
 885 890 895
 Thr Leu Val Leu Ser Val Ile Gly Ser Val Val Ile Val Asn Phe Ile
 900 905 910
 Leu Ser Thr Thr Ile
 915

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Ala Asn Ser Ser Val Pro Pro Ile Thr Val Asn Ala Lys Met Asn Lys
 1 5 10 15
 Asp Val Asn Ser Phe
 20

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Asp Asn Gly Ala Gly Ala Asp Ser Phe Lys Asn Asp Gly Val Tyr Ser
 1 5 10 15
 Arg Tyr Phe Thr Ala Tyr Thr Glu Asn Gly Arg Tyr Ser Leu Lys
 20 25 30

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Val	Arg	Ala	His	Gly	Gly	Ala	Asn	Thr	Ala	Arg	Leu	Lys	Leu	Arg	Pro
1				5					10					15	
Pro	Leu	Asn	Arg	Ala	Ala	Tyr	Ile								20

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Ser	Leu	Pro	Leu	Pro	Asp	Gln	Tyr	Pro	Pro	Ser	Gln	Ile	Thr	Asp	Leu
1					5				10					15	
Asp	Ala	Thr	Val	His	Glu	Asp	Lys	Ile	Ile	Leu	Thr	Trp	Thr	Ala	Pro
								20	25				30		
Gly	Asp	Asn	Phe	Asp	Val	Gly	Lys								35
									40						

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Tyr	Asn	Glu	Asp	Gln	Pro	Phe	Tyr	Arg	Ala	Lys	Ser	Lys	Lys	Ile	Glu
1				5					10					15	
Ala	Thr	Arg	Cys												20

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Leu Ser Arg Ala Cys Arg Ile Asp Ser Thr Thr Lys Leu Tyr Gly Lys
 1 5 10 15
 Asp Cys Gln Phe Phe Pro Asp Lys
 20

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Lys Ser Ser Asp Glu Arg Asn Thr Leu Met Ala Gly Leu Pro Thr Tyr
 1 5 10 15
 Pro Leu Gly Gly
 20

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Glu Ile Asp Glu Asp Thr Gln Thr Thr Leu Glu Asp Phe Ser Arg
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Asp Tyr Lys Asp Asp Asp Asp Lys
 1 5

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Asn Met His Thr Glu His
1 5 10 15
His His His His His
20

We Claim:

1. A purified polynucleotide or fragment thereof derived from a CS193 gene, wherein said polynucleotide is capable of selectively hybridizing to the nucleic acid of said CS193 gene and has at least 50% identity to a sequence selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.

10 2. The purified polynucleotide of claim 1, wherein said polynucleotide is produced by recombinant techniques.

15 3. The purified polynucleotide of claim 1, wherein said polynucleotide is produced by synthetic techniques.

4. The purified polynucleotide of claim 1, wherein said polynucleotide comprises a sequence encoding at least one CS193 epitope.

20 5. A recombinant expression system comprising a nucleic acid sequence that includes an open reading frame derived from CS193 operably linked to a control sequence compatible with a desired host, wherein said nucleic acid sequence has at least 50% identity to a sequence selected from the group consisting of SEQUENCE ID NOS 1-18 and fragments or complements thereof.

25 6. A cell transfected with the recombinant expression system of claim 5.

7. A CS193 polypeptide having at least 60% identity with an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

30 8. The polypeptide of claim 7, wherein said polypeptide is produced by recombinant techniques.

35 9. The polypeptide of claim 7, wherein said polypeptide is produced by synthetic techniques.

10. An antibody which specifically binds to at least one CS193 epitope, wherein said CS193 epitope is derived from an amino acid sequence having at least 50% identity to an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

5

11. A cell transfected with a nucleic acid sequence encoding at least one CS193 epitope, wherein said nucleic acid sequence is selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.

10 12. A method for producing a polypeptide comprising at least one CS193 epitope, said method comprising incubating host cells that have been transfected with an expression vector containing a polynucleotide sequence encoding a polypeptide, wherein said polypeptide comprises an amino acid sequence having at least 60% identity with an amino acid sequence selected from the group consisting of 15 SEQUENCE ID NOS 41-49, and fragments thereof.

20 13. A method for producing antibodies which specifically bind to CS193 antigen, said method comprising administering to an individual an isolated immunogenic polypeptide or fragment thereof in an amount sufficient to elicit an immune response, wherein said immunogenic polypeptide comprises at least one CS193 epitope and has at least 50% identity with a sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

25 14. A method for producing antibodies which specifically bind to CS193 antigen, said method comprising administering to an individual a plasmid comprising a polynucleotide sequence which encodes at least one CS193 epitope derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

30 15. A composition of matter comprising a CS193 polynucleotide or fragment thereof, wherein said polynucleotide has at least 50% identity with a polynucleotide selected from the group consisting of SEQUENCE ID NOS 1-18, and fragments or complements thereof.

35 16. A composition of matter comprising a polypeptide containing at least one CS193 epitope, wherein said polypeptide has at least 60% identity with a sequence

selected from the group consisting of SEQUENCE ID NOS 41-49, and fragments thereof.

17. A gene, or a fragment thereof, which codes for a CS193 protein
5 comprising an amino acid sequence that has at least 60% identity with SEQUENCE ID
NO 41.

18. A gene or fragment thereof comprising DNA having at least 50%
identity with SEQUENCE ID NO 16, SEQUENCE ID NO 17, or SEQUENCE ID NO
10 18.

REAGENTS AND METHODS USEFUL FOR DETECTING
DISEASES OF THE GASTROINTESTINAL TRACT

5

Abstract of the Disclosure

10 A set of contiguous and partially overlapping cDNA sequences and polypeptides encoded thereby, designated as CS193 and transcribed from GI tract tissue, are described. These sequences are useful for the detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating, or determining the predisposition of an individual to diseases and conditions of the GI tract, such as GI tract cancer. Also
15 provided are antibodies which specifically bind to CS193-encoded polypeptide or protein, and agonists or inhibitors which prevent action of the tissue-specific CS193 polypeptide, which molecules are useful for the therapeutic treatment of GI tract diseases, tumors or metastases.



Figure 1-A

>2767646	GCCAGGAATA ACTAGAGAGG AACAAATGGGG TTATTCAAGAG GTTTTGT
>774134	CTAGAGAGG AACAAATGGGG TTATTCAAGAG GTTTTGT
>774134IH	CTAGAGAGG AACAAATGGGG TTATTCAAGAG GTTTTGT
Consensus	GCCAGGAATA ACTAGAGAGG AACAAATGGGG TTATTCAAGAG GTTTTGT
>2767646	CCTCTTAGTT CTGTGCCTGC TGCACCAGTC AAATACTTCC TTCATTAAGC
>774134	CCTCTTAGTT CTGTGCCTGC TGCACCAGTC AAATACTTCC TTCATTAAGC
>774134IH	CCTCTTAGTT CTGTGCCTGC TGCACCAGTC AAATACTTCC TTCATTAAGC
Consensus	CCTCTTAGTT CTGTGCCTGC TGCACCAGTC AAATACTTCC TTCATTAAGC
>2767646	TGAATAATAA TGGCTTGAA GATATTGTCA TTGTTATAGA TCCTAGTGTG
>774134	TGAATAATAA TGGCTTGAA GATATTGTCA TTGTTATAGA TCCTAGTGTG
>774134IH	TGAATAATAA TGGCTTGAA GATATTGTCA TTGTTATAGA TCCTAGTGTG
Consensus	TGAATAATAA TGGCTTGAA GATATTGTCA TTGTTATAGA TCCTAGTGTG
>2767646	CCAGAACATG AAAAAATAAT TGAACAAATA GAGGATATGG TGACTACAGC
>774134	CCAGAACATG AAAAAATAAT TGAACAAATA GAGGATATGG TGACTACAGC
>774134IH	CCAGAACATG AAAAAATAAT TGAACAAATA GAGGATATGG TGACTACAGC
Consensus	CCAGAACATG AAAAAATAAT TGAACAAATA GAGGATATGG TGACTACAGC
>2767646	TTCTACGTAC CTGTTGAAG CCACAGAAAA AAGATTTTT T
>774134	TTCTACGTAC CTGTTGAAG CCACAGAAAA
>774134IH	TTCTACGTAC CTGTTGAAG CCACAGAAAA AAGATTTTT TTCAAAATG
Consensus	TTCTACGTAC CTGTTGAAG CCACAGAAAA AAGATTTTT TTCAAAATG
>774134IH	TATCTATATT AATTCTGAG AATTGGAAGG AAAATCCTCA GTACAAAAGG
Consensus	TATCTATATT AATTCTGAG AATTGGAAGG AAAATCCTCA GTACAAAAGG
>774134IH	CCAAAACATG AAAACCATAA ACATGCTGAT GTTATAGTTG CACCACCTAC
Consensus	CCAAAACATG AAAACCATAA ACATGCTGAT GTTATAGTTG CACCACCTAC
>774134IH	ACTCCCAGGT AGAGATGAAC CATACACCAA GCAGTTCAC AATGTGGAG
Consensus	ACTCCCAGGT AGAGATGAAC CATACACCAA GCAGTTCAC AATGTGGAG
>774134IH	AGAAAGCGA ATACATTAC TTCACCCCTG ACCTTCTACT TGAAAAAAAAA
Consensus	AGAAAGCGA ATACATTAC TTCACCCCTG ACCTTCTACT TGAAAAAAAAA
>774134IH	CAAAATGAAT ATGGACCACC AGGCAAACCTG TTTGTCCATG AGTGGGCTCA
Consensus	CAAAATGAAT ATGGACCACC AGGCAAACCTG TTTGTCCATG AGTGGGCTCA
>774134IH	CCTCCGGTGG GGAGTGTITG ATGAGTACAA TGAAGATCAG CCTTTCTACC
Consensus	CCTCCGGTGG GGAGTGTITG ATGAGTACAA TGAAGATCAG CCTTTCTACC
>774134IH	GTGCTAAGTC AAAAAAAATC GAAGCAACAA GGTGTTCCGC AGGTATCTCT
Consensus	GTGCTAAGTC AAAAAAAATC GAAGCAACAA GGTGTTCCGC AGGTATCTCT
>774134IH	GGTAGAAATA GAGTTTATAA GTGTCAGGA GGCAGCTGTC TTAGTAGAGC
Consensus	GGTAGAAATA GAGTTTATAA GTGTCAGGA GGCAGCTGTC TTAGTAGAGC

Figure 1-B



>774134IH Consensus	ATGCAGAATT GATTCTACAA CAAAAGTGA TGGAAAAGAT TGTCAATTCT ATGCAGAATT GATTCTACAA CAAAAGTGA TGGAAAAGAT TGTCAATTCT
>774134IH Consensus	TTCCTGATAA AGTACAAACA GAAAAAGCAT CCATAATGTT TATGCAAAGT TTCCTGATAA AGTACAAACA GAAAAAGCAT CCATAATGTT TATGCAAAGT
>774134IH >775437 Consensus	ATTGATTCTG TTGTTGAATT TTGTAACGAA AAAACCCATA ATCAAGAAC TT NTGTAACGAA AAAACCCATA ATCAAGAAC ATTGATTCTG TTGTTGAATT TTGTAACGAA AAAACCCATA ATCAAGAAC
>774134IH >775437 Consensus	TCCAAGCCTA CAAAACATAA AGTCAATT TAGAAGTACA TGGGAGGTGA TCCAAGCCTA CAAAACATAA AGTCAATT TAGAAGTACA TGGGAGGTGA TCCAAGCCTA CAAAACATAA AGTCAATT TAGAAGTACA TGGGAGGTGA
>774134IH >775437 Consensus	TTAGCAATT TGAGGATTAA AAAAACACCA TACCCATGGT GACACCACCT TTAGCAATT TGAGGATTAA AAAAACACCA TACCCATGGT GACACCACCT TTAGCAATT TGAGGATTAA AAAAACACCA TACCCATGGT GACACCACCT
>774134IH >775437 Consensus	CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT CCTCCACCTG TCTTCTCATT GCTGAAGATC AGTCAAAGAA TTGTGTGCTT
>774134IH >775437 >1281329 Consensus	AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAATCGAA AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAATCGAA TGGGGGGG TAAGGACCGC CTAATCGAA AGTTCTTGAT AAGTCTGGAA GCATGGGGGG TAAGGACCGC CTAATCGAA
>774134IH >1281329 Consensus	TGAATCAAGC ACCAAAACAT TTCCCTGCTGC AGACTGTTGA AAATGGATCC TGAATCAAGC ACCAAAACAT TTCCCTGCTGC AGACTGTTGA AAATGGATCC TGAATCAAGC AGCAAAACAT TTCCCTGCTGC AGACTGTTGA AAATGGATCC
>774134IH >1281329 Consensus	TGGGTGGGGA TGGTTCACTT TGATAGTACT GCCACTATTG TAAATAAGCT TGGGTGGGGA TGGTTCACTT TGATAGTACT GCCACTATTG TAAATAAGCT TGGGTGGGGA TGGTTCACTT TGATAGTACT GCCACTATTG TAAATAAGCT
>774134IH >1281329 Consensus	AATCCAAATA AAAAGCAGTG ATGAAAGAAA CACACTCATG GCAGGATTAC AATCCAAATA AAAAGCAGTG ATGAAAGAAA CACACTCATG GCAGGATTAC AATCCAAATA AAAAGCAGTG ATGAAAGAAA CACACTCATG GCAGGATTAC
>774134IH >1281329 >1628677 Consensus	CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA CTTCCATCT GCTCTGGAAT TAAATATGCA CTACATATCC TCTGGGAGGA ACTTCCATCT GCTCTGGAAT TAAATATGCA
>774134IH >1281329 >1628677 Consensus	TTTCAGGTGA TTGGAGAGCT ACATTCCAA CTCGATGGAT CCGAAGTACT TTTCAGGTGA TTTCAGGTGA TTGGAGAGCT ACATTCCAA CTCGATGGAT CCGAAGTACT TTTCAGGTGA TTGGAGAGCT ACATTCCAA CTCGATGGAT CCGAAGTACT



Figure 1-C

>774134IH >1628677 Consensus	GCTGCTGACT GATGGGGAGG ATAACACTGC AAGTTCTTGT ATTGATGAAG GCTGCTGACT GATGGGGAGG ATAACACTGC AAGTTCTTGT ATTGATGAAG GCTGCTGACT GATGGGGAGG ATAACACTGC AAGTTCTTGT ATTGATGAAG
>774134IH >1628677 Consensus	TGAAACAAAG TGGGCCATT GTTCATTTA TTGCTTGGG AAGAGCTGCT TGAAACAAAG TGGGCCATT GTTCATTTA TTGCTTGGG AAGAGCTGCT TGAAACAAAG TGGGCCATT GTTCATTTA TTGCTTGGG AAGAGCTGCT
>774134IH >1628677 Consensus	GATGAAGCAG TAATAGAGAT GAGCAAGATA ACAGGAGGAA GTCATTTTA GATGAAGCAG TAATAGAGAT GAGCAAGATA ACAGGAG GATGAAGCAG TAATAGAGAT GAGCAAGATA ACAGGAGGAA GTCATTTTA
>774134IH Consensus	TGTTTCAGAT GAAGCTCAGA ACAATGGCCT CATTGATGCT TTTGGGGCTC TGTTTCAGAT GAAGCTCAGA ACAATGGCCT CATTGATGCT TTTGGGGCTC
>774134IH Consensus	TTACATCAGG AAATACTGAT CTCTCCAGA AGTCCTTCA GCTGAAAAGT TTACATCAGG AAATACTGAT CTCTCCAGA AGTCCTTCA GCTGAAAAGT
>774134IH >1286372 Consensus	AAGGGATTAA CACTGAATAG TAATGCCTGG ATGAACGACA CTGTCATAAT AAGGGATTAA CACTGAATAG TAATGCCTGG ATGAACGACA CTGTCATAAT
>774134IH >1286372 Consensus	TGATAGTACA CTGGAAAGG ACACGTTCTT TCTCATCAC A TGGAACAGTC TGATAGTACA CTGGAAAGG NCACGTTCTT TCTCATCAC A TGGAACAGTC TGATAGTACA CTGGAAAGG ACACGTTCTT TCTCATCAC A TGGAACAGTC
>774134IH >1286372 Consensus	TGCCTCCAG TATTTCTCTC TGGGATCCC GTGGAACAAT AATGGAAAAT TGCCTCCAG TATTTCTCTC TGGGATCCC GTGGAACAAT AATGGAAAAT TGCCTCCAG TATTTCTCTC TGGGATCCC GTGGAACAAT AATGGAAAAT
>774134IH >1286372 Consensus	TTCACAGTGG ATGCAACTTC CAAAATGGCC TATCTCAGTA TTCCAGGAAC TTCACAGTGG ATGCAACTTC CAAAATGGCC TATCTCAGTA TTCCAGGAAC TTCACAGTGG ATGCAACTTC CAAAATGGCC TATCTCAGTA TTCCAGGAAC
>774134IH >1286372 Consensus	TGCAAAGGTG GGCACTTGGG CATAAACATCT TCAAGCCAAA GCGAACCCAG TGCAAAGGTG GGCACTTGGG CATAAACATCT TCAAGCCAAA GCGAACCC TGCAAAGGTG GGCACTTGGG CATAAACATCT TCAAGCCAAA GCGAACCCAG
>774134IH >774419 >774419IH Consensus	AAACATTAAC TATTACAGTA ACTTCTCGAG CAGCAAATTC TTCTGTGCCT GCAAATTC TTCTGTGCCT GCAAATTC TTCTGTGCCT AAACATTAAC TATTACAGTA ACTTCTCGAG CAGCAAATTC TTCTGTGCCT
>774134IH >774419 >774419IH Consensus	CCAATCACAG TGAATGCTAA AATGAATAAG GACGTAACAA GTTCCCCAG CCAATCACAG TGAATGCTAA AATGAATAAG GACGTAACAA GTTCCCCAG CCAATCACAG TGAATGCTAA AATGAATAAG GACGTAACAA GTTCCCCAG CCAATCACAG TGAATGCTAA AATGAATAAG GACGTAACAA GTTCCCCAG

Figure 1-D



>774134IH	CCCAATGATT GTTTACGCAG AAATTCTACA AGGATATGTA CCTGTTCTTG
>774419	CCCAATGATT GTTTACGCAG AAATTCTACA AGGATATGTA CCTGTTCTTG
>774419IH	CCCAATGATT GTTTACGCAG AAATTCTACA AGGATATGTA CCTGTTCTTG
Consensus	CCCAATGATT GTTTACGCAG AAATTCTACA AGGATATGTA CCTGTTCTTG
>774134IH	GAGCCAATGT GACTGCTTTC ATTGAATCAC AGAATGGACA TACAGAAGTT
>774419	GAGCCAATGT GACTGCTTTC ATTGAATCAC AGAATGGACA TACAGAAGTT
>774419IH	GAGCCAATGT GACTGCTTTC ATTGAATCAC AGAATGGACA TACAGAAGTT
Consensus	GAGCCAATGT GACTGCTTTC ATTGAATCAC AGAATGGACA TACAGAAGTT
>774134IH	TTGGAACCTTT TGGATAATGG TGCAAGCGCT GATTCTTCA AGAATGATGG
>774419	TTGGAACCTTT TGGATAATGG TGCAAGCGCT GATTCTTCA AGAATGATGG
>774419IH	TTGGAACCTTT TGGATAATGG TGCAAGCGCT GATTCTTCA AGAATGATGG
>3233118	TTGGAACCTTT TGGATAATGG TGCAAGCGCT GATTCTTCA AGAATGATGG
Consensus	TTGGAACCTTT TGGATAATGG TGCAAGCGCT GATTCTTCA AGAATGATGG
>774134IH	AGTCTACTCC AGGTATTTTA CAGCATATAC AGAAAATGGC AGATATAGCT
>774419	AGTCTACTCC AGGTATTTTA CAG
>774419IH	AGTCTACTCC AGGTATTTTA CAGCATATAC AGAAAATGGC AGATATAGCT
>3233118	AGTCTACTCC AGGTATTTTA CAGCATATAC AGAAAATGGC AGATATAGCT
Consensus	AGTCTACTCC AGGTATTTTA CAGCATATAC AGAAAATGGC AGATATAGCT
>774134IH	AAAAAGTTCG GGCTCATGGA GGAGCAAACA CTGCCAGGCT AAAATTACGG
>774419IH	AAAAAGTTCG GGCTCATGGA GGAGCAAACA CTGCCAGGCT AAAATTACGG
>3233118	AAAAAGTTCG GGCTCATGGA GGAGCAAACA CTGCCAGGCT AAAATTACGG
Consensus	AAAAAGTTCG GGCTCATGGA GGAGCAAACA CTGCCAGGCT AAAATTACGG
>774134IH	CCTCCACTGA ATAGAGCCGC GTACATACCA GGCTGGTAG TGAACGGGA
>774419IH	CCTCCACTGA ATAGAGCCGC GTACATACCA GGCTGGTAG TGAACGGGA
>3233118	CCTCCACTGA ATAGAGCCGC GTACATACCA GGCTGGTAG TGAACGGGA
Consensus	CCTCCACTGA ATAGAGCCGC GTACATACCA GGCTGGTAG TGAACGGGA
>774134IH	AATTGAAGCA AACCCGCCAA GACCTGAAAT TGATGAGGAT ACTCAGACCA
>774419IH	AATTGAAGCA AACCCGCCAA GACCTGAAAT TGATGAGGAT ACTCAGACCA
>3233118	AATTGAAGCA AACCCGCCAA GACCTGAAAT TGATGAGGAT ACTCAGACCA
>2733923	CCCGCCAA GACCTGAAAT TGATGAGGAT ACTCAGACCA
Consensus	AATTGAAGCA AACCCGCCAA GACCTGAAAT TGATGAGGAT ACTCAGACCA
>774134IH	CCTTGAGGA TTTCAGCCGA ACAGCATCCG GAGGTGCATT TGTGGTATCA
>774419IH	CCTTGAGGA TTTCAGCCGA ACAGCATCCG GAGGTGCATT TGTGGTATCA
>3233118	CCTTGAGGA T
>2733923	CCTTGAGGA TTTCAGCCGA ACAGCATCCG GAGGTGCATT TGTGGTATCA
Consensus	CCTTGAGGA TTTCAGCCGA ACAGCATCCG GAGGTGCATT TGTGGTATCA
>774134IH	CAAGTCCCAA GCCTTCCCTT GCCTGACCAA TACCCACCAA GTCAAATCAC
>774419IH	CAAGTCCCAA GCCTTCCCTT GCCTGACCAA TACCCACCAA GTCAAATCAC
>2733923	CAAGTCCCAA GCCTTCCCTT GCCTGACCAA TACCCACCAA GTCAAATCAC
>906,605	CCAA TACCCACCAA GTCAAATNAC
Consensus	CAAGTCCCAA GCCTTCCCTT GCCTGACCAA TACCCACCAA GTCAAATCAC



Figure 1-E

>774134IH	AGACCTTGAT GCCACAGTTC ATGAGGATAA GATTATTCTT ACATGGACAG
>774419IH	AGACCTTGAT GCCACAGTTC ATGAGGATAA GATTATTCTT ACATGGACAG
>2733923	AGACCTTGAT GCCACAGTTC ATGAGGATAA GATTATTCTT ACATGGACAG
>906605	AGACCTTGAT GCCACAGTTN ATGAGGATAA GATTATTCTT ACATGGACAG
Consensus	AGACCTTGAT GCCACAGTTC ATGAGGATAA GATTATTCTT ACATGGACAG
>774134IH	CACCAAGGAGA TAATTTTGAT GTTGGAAAAG TTCAACGTTA TATCATAAAGA
>774419IH	CACCAAGGAGA TAATTTTGAT GTTGGAAAAG TTCAACGTTA TATCATAAAGA
>2733923	CACCAAGGAGA TAATTTTGAT GTTGGAAAAG TTCAACGTTA TATCA
>906605	CACCAAGGAGA TAATTTTGAT GTTGGAAAAG TTCAACGTTA TATCATAAAGA
Consensus	CACCAAGGAGA TAATTTTGAT GTTGGAAAAG TTCAACGTTA TATCATAAAGA
>774134IH	ATAAGTGCAA GTATTCTTGA TCTAAGAGAC AGTTTTGATG ATGCTCTTCA
>774419IH	ATAAGTGCAA GTATTCTTGA TCTAAGAGAC AGTTTTGATG ATGCTCTTCA
>906605	ATAAGTGCAA GTATTCTTGA TCTAAGAGAC AGTTTTGATG ATGCTCTTCA
>2771475	AA CTATTCTTGA TCTAAGAGAC AGTTTTGATG ATGCTCTTCA
Consensus	ATAAGTGCAA GTATTCTTGA TCTAAGAGAC AGTTTTGATG ATGCTCTTCA
>774134IH	AGTAAATACT ACTGATCTGT CACCAAAAGGA GGCCAACCTCC AAGGAAAGCT
>774419IH	AGTAAATACT ACTGATCTGT CACCAAAAGGA GGCCAACCTCC AAGGAAAGCT
>906605	AGTAAATACT ACTGATCTGT CACCAAAAGGA GGCCAACCTCC AAGGAAAGCT
>2771475	AGTAAATACT ACTGATCTGT CACCAAAAGGA GGCCAACCTCC AAGGAAAGCT
Consensus	AGTAAATACT ACTGATCTGT CACCAAAAGGA GGCCAACCTCC AAGGAAAGCT
>774134IH	TTGCATTTAA ACCAGAAAAT ATCTCAGAAC AAAATGCAAC CCACATATTG
>774419IH	TTGCATTTAA ACCAGAAAAT ATCTCAGAAC AAAATGCAAC CCACATATTG
>906605	TTGCATTTAA ACCAGAAAAT ATCTCAGAAC AAAATGCAAC CCACATATTG
>2771475	TTGCATTTAA ACCAGAAAAT ATCTCAGAAC AAAATGCAAC CCACATATTG
>1803247	AN ANAATGCAAC CCACATATTG
Consensus	TTGCATTTAA ACCAGAAAAT ATCTCAGAAC AAAATGCAAC CCACATATTG
>774134IH	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
>774419IH	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
>906605	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
>2771475	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
>1803247	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
Consensus	ATTGCCATTA AAAGTATAGA TAAAAGCAAT TTGACATCAA AAGTATCCAA
>774134IH	CATTGCACAA GTAACTTTGT TTATCCCTCA AGCAAATCCT GATGACATTG
>774419IH	CATTGCACAA GTAACTTTGT TTATCCCTCA AGCAAATCCT GATGACATTG
>2771475	CATTGCACAA GTAACTTTGT TTATCCCTCA AGCAAATCCT GATGACATTG
>1803247	CATTGCACAA GTAACTTTGT TTATCCCTCA AGCAAATCCT GATGACATTG
Consensus	CATTGCACAA GTAACTTTGT TTATCCCTCA AGCAAATCCT GATGACATTG
>774134IH	ATCCTACTCC TACTCCTACT CCTACTCCTG ATAAAAGTCA TAATTCTGGA
>774419IH	ATCCTACTCC TACTCCTACT CCTACTCCTG ATAAAAGTCA TAATTCTGGA
>1803247	ATCCTACTCC TACTCCTACT CCTACTCCTG ATAAAAGTCA TAATTCTGGA
>1737526	CTCCTACT CCTACTCCTG ATAAAAGTCA TAATTCTGGA
Consensus	ATCCTACTCC TACTCCTACT CCTACTCCTG ATAAAAGTCA TAATTCTGGA

Figure 1-F



	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>774419IH	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>1803247	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>1737526	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>2792957	TCTG TTGTAATTGT
Consensus	GTTAATATTT CTACGCTGGT ATTGTCTGTG ATTGGGTCTG TTGTAATTGT
>774134IH	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAAATCT
>774419IH	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAAATCT
>1737526	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAAATCT
>2792957	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAAATCT
>1226186	G AAAAAAAATCT
Consensus	TAACTTTATT TTAAGTACCA CCATTTGAAC CTTAACGAAG AAAAAAAATCT
>774134IH	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTAAAG
>774419IH	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTAAAG
>1737526	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTAAAG
>2792957	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTNAAG
>1226186	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTAAAG
Consensus	TCAAGTAGAC CTAGAAGAGA GTTTTAAAAA ACAAAACAAT GTAAGTAAAG
>774134IH	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
>774419IH	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
>1737526	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
>2792957	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
>1226186	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
Consensus	GATATTTCTG AATCTTAAAA TTCATCCCAT GTGTGATCAT AAACTCATAA
>774134IH	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAACACT
>774419IH	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAACACT
>2792957	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAACACT
>1226186	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAACACT
Consensus	AAATAATTTT AAGATGTCGG AAAAGGATAC TTTGATTAAA TAAAAACACT
>774134IH	CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>774419IH	CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>2792957	CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>1226186	CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
Consensus	CATGGATATG TAAAAACTGT CAAGATTAAA ATTTAATAGT TTCATTTATT
>774134IH	TGTTATTTA TTTGTAAGAA ATAGTGATGA ACAAAAGATCC TTTTCATAC
>774419IH	TGTTATTTA TTTGTAAGAA ATAGTGATGA ACAAAAGATCC TTTTCATAC
>2792957	TGTTATTTA TTTGTAAGAN ATAGTGATGA ACAAAAGA
>1226186	TGTTATTTA TTTGTAAG
Consensus	TGTTATTTA TTTGTAAGAA ATAGTGATGA ACAAAAGATCC TTTTCATAC
>774134IH	TGAT
>774419IH	TGATACCTGG TTGTATATTA TTTGATGCAA CAGTTTCTG AAATGATATT
Consensus	TGATACCTGG TTGTATATTA TTTGATGCAA CAGTTTCTG AAATGATATT
>774419IH	TCAAATTGCA TCAAGAAATT AAAATCATCT ATCTGAGTAG TCAAAATACA
Consensus	TCAAATTGCA TCAAGAAATT AAAATCATCT ATCTGAGTAG TCAAAATACA



Figure 1-G

>774419IH
Consensus

AGTAAAGGAG AGCAAATAAA CAACATTTGG A
AGTAAAGGAG AGCAAATAAA CAACATTTGG A

Figure 2

